Comparison between Unitary and Causal Approaches to Transform Coding of Vectorial Signals

David Mary[1], Dirk T.M. Slock[1],

[1]Institut EURÉCOM *
2229, Route des Crêtes, BP 193, 06904 Sophia Antipolis, France
David.Mary@eurecom.fr, Dirk.Slock@eurecom.fr

Résumé – Dans le cadre du codage par transformée, nous comparons en terme de gain de codage deux approches : l'approche causale optimale (factorisation LDU, Lower-Diagonal-Upper) et l'approche unitaire optimale (Karhunen-Loeve Transform, KLT). Les deux transformations produisent le même gain quand elles sont basées sur le signal original. Le but de ce travail est de comparer les comportements de ces transformations quand elles sont perturbées. Cette comparaison est menée pour deux cas courants de perturbation. Le premier cas est celui du bruit de quantification lié à l'adaptation en boucle fermée des transformations. Nous montrons que dans le cas causal, un retour de bruit dégrade le gain de codage. Le deuxième cas de perturbation est le bruit d'estimation lié au nombre limité de données disponibles. Ce cas est traité sous l'hypothèse de vecteurs indépendants identiquement distribués. L'espérance du gain de codage dans le cas causal apparaît comme indépendante de la matrice de covariance R, alors que dans le cas unitaire cette espérance dépend fortement de la distribution des valeurs propres de R.

Abstract – In a transform coding framework we compare, in terms of coding gain, the optimal causal approach (LDU, Lower-Diagonal-Upper) to the optimal unitary approach (Karhunen-Loeve Transform, KLT). Both transforms are known to yield the same gain when they are based on the original signal. The purpose of this paper is to compare the behaviour of the two transformations when the ideal transform coding scheme gets perturbed. This comparison is made in two usual cases of perturbation. The first perturbation we consider is due to the quantization noise (occuring when the transformations are backward adapted). We show that under high resolution assumption, a quantization noise feedback occurs in the causal scheme, which decreases the coding gain. The second perturbation considered in this work is the estimation noise, due to a finite number of available data, under the assumption of independent identically distributed vectors. The expectation of the coding gain in the causal case is shown to be independent of the covariance matrix R, whereas in the unitary case, the expectation of the coding gain appears strongly sensitive to the distribution of the eigenvalues of R.

1 Introduction

Consider a vectorial signal whose samples are X_i. In the transform coding framework, a matricial transformation is applied to each vector X_i to produce a vector Y_i. Each component of Y_i is then independently quantized using a scalar quantizer Q_i. The optimal causal transform has recently been shown [3, 2] to correspond to an LDU triangular factorization of the autocorrelation matrix of the signal. The optimal unitary transform is the well-known KLT. When these transformations are adapted on the original signal, the causal (LDU) and unitary transformations are both optimal. The aim of this paper is to compare the behaviour of the causal and unitary approaches when the ideal scheme gets perturbed. The optimal causal transform and the coding gains for the two approaches are reviewed in the second part.

We consider then separately two types of perturbations. The first type is due to the quantization noise occuring in backward adapted schemes . Though both transformation were compared in [3], the performance of the optimal causal transformation in terms of coding gain was not investigated when the transformation is adapted on the quantized signal. Besides, it was shown in [4] that under some assumptions the backward adapted KLT converges to the optimal KLT. The comparison between the two approaches in the presence of quantization noise is made in the third part.

The second type of perturbation we consider in this paper is the estimation noise : the transformation is computed on the basis of a finite amount data, whose statistics are thus not perfectly known. We suppose in this case independent identically distributed X_i (which is for example the case when the sampling period is low in comparison with the typical time of correlation of the X_i.)

2 Coding Gains without Perturbation

The optimal causal transform may be described as follows. Let us consider the generalization of the classical DPCM coding scheme applied to a vector $X = [x_1 ... x_N]^T$. A matrix transformation L is applied to the vector X : $Y = LX = X - \overline{L}X$, where $\overline{L}X$ is the reference vector. The difference vector $Y = [y_1 ... y_N]^T$ is then quantized using a

*Eurécom's research is partially supported by its industrial partners: Ascom, Swisscom, Thales Communications, STMicroelectronics, CEGETEL, Motorola, France Télécom, Bouygues Telecom, Hitachi Europe Ltd. and Texas Instruments.

set $\mathbf{Q}$ of quantizers Q_i. The output X^q is $Y^q + \overline{L}X$. Note that the reconstruction error $\tilde{X}$ equals the quantization error $\tilde{Y}$:

$$\tilde{X} = X - X^q = X - (Y^q + \overline{L}X) = X - \overline{L}X - Y^q = Y - Y^q = \tilde{Y}, \quad (1)$$

as in the unitary case. The constraint imposed on the transformation here is causality, which imposes a lower triangular structure. The unitary aspect of the transform appears in the unicity of the main diagonal ($\overline{L} = I - L$ is hence strictly lower triangular and represents the degrees of freedom of the transformation). The notion of causality could be generalized by working with the permuted components of X and Y, which gives $\mathcal{P}Y = L\ \mathcal{P}X$ or $Y = (\mathcal{P}^T L \mathcal{P})X$, where $\mathcal{P}$ is a permutation matrix. The coding gain for a transformation L is

$$G_L = \frac{E\|\tilde{X}\|^2_{(I)}}{E\|\tilde{X}\|^2_{(L)}} = \frac{E\|\tilde{X}\|^2_{(I)}}{E\|\tilde{Y}\|^2_{(L)}}, \quad (2)$$

where I is the identity matrix (which corresponds to the absence of transformation), and the notation $\|\tilde{X}\|^2_{(T)}$ denotes the variance of the quantization error on the vector X, obtained for a transformation T. The second equality in (2) follows from the equality (1), as in the unitary case. Thus, the coding gains for both transforms are equal and can be derived as follows. A quantizer Q_i introduces an independent white noise $\tilde{y_i}$ on the component y_i, of variance $\sigma^2_{\tilde{y}_i} = c\, 2^{-2R_i}\sigma^2_{y_i}$, where R_i is the number of bits assigned to the quantizer Q_i, and c is a constant depending on the probability density function of the signal to be quantized (one should assume a Gaussian distribution, linear transform invariant).
For a given L, the optimal bit assignment has to minimize $E\|\tilde{Y}\|^2_{(L)} = \sum_{i=1}^N \sigma^2_{y_i} c 2^{-2R_i}$ under the constraint $\sum_{i=1}^N R_i = NR$, where R is the average number of bits assigned to the N quantizers Q_i. Using well-known techniques, and making abstraction of the fact that the R_i are integer and non negative, one shows that

$$\sigma^2_{\tilde{y}_i} = c\, 2^{-2R_i}\sigma^2_{y_i} = c\, 2^{-2R}\left(\prod_{i=1}^N \sigma^2_{y_i}\right)^{\frac{1}{N}} . \quad (3)$$

Note that the optimal quantization error variances $\sigma^2_{\tilde{y}_i}$ are equal (independent of i).
As shown in [2],the optimal L (in terms of coding gain) is such that

$$L R_{XX} L^T = R_{YY} = D = diag\{\sigma^2_{y_1}, ...\sigma^2_{y_N}\}, \quad (4)$$

where $diag\{...\}$ represents a diagonal matrix whose elements are $\sigma^2_{y_i}$. In other words, the components y_i are the prediction errors of x_i with respect to the past values of X, the $X_{1:i-1}$, and the optimal coefficients $-L_{i,1:i-1}$ are the optimal prediction coefficients. Since each prediction error y_i is orthogonal to the subspaces generated by the $X_{1:i-1}$, the y_i are orthogonal, and D is diagonal. It follows that

$$R_{XX} = L^{-1} R_{YY} L^{-T}, \quad (5)$$

which represents the LDU factorization of R_{XX}. Referring to (2), the coding gain without perturbation for the optimal causal transform can be written as

$$G_L^{(0)} = \left(\frac{\det\left[diag(R_{XX})\right]}{\det\left[diag(L R_{XX} L^T)\right]}\right)^{\frac{1}{N}} \quad (6)$$

where $diag(R)$ denotes here the diagonal matrix that corresponds to the diagonal of the matrix R.
Since the diagonalizing transformation matrix L is unimodular, $\det(diag(R_{YY})) = \det(R_{XX}) = \det\Lambda$, where Λ is the eigenvalue matrix of R_{XX}. Thus

$$\begin{aligned} G_L^{(0)} &= \left(\frac{\det[diag(R_{XX})]}{\det[diag(L R_{XX} L^T)]}\right)^{\frac{1}{N}} \\ &= \left(\frac{\det[diag(R_{XX})]}{\det\Lambda}\right)^{\frac{1}{N}} = G^{(0)}_{KLT} . \end{aligned} \quad (7)$$

3 Quantization Effects on the Coding Gains

Let us now inspect the case where the tranformations are not based on the original signal but on its quantized version. Thus, the asymptotic (with respect to the amount of data) estimated covariance matrix of the quantized data is $R_{XX} + \Delta R$, where ΔR denotes the perturbation term due to the quantization.

3.1 Causal Approach

In this case, the difference vector of (1) becomes

$$Y = X - \overline{L}X^q = X - \overline{L}(X - \tilde{X}) = LX + \overline{L}\tilde{Y} . \quad (8)$$

Y now not only contains the prediction error LX of X, but also the quantization error $\tilde{Y}$ filtered by the optimal predictor $\overline{L}$. In this case again, the optimal bit assignment has to minimize the sum of the $\sigma^2_{\tilde{y}_i}$. It follows that the variances of the quantization noises are $\sigma^2_{\tilde{y}_i} = c2^{-2R}(\prod_{i=1}^N \sigma^2_{y_i})^{\frac{1}{N}} = \sigma^2_{\tilde{y}_1}$, independent of i. The autocorrelation matrix of the noise is hence $R_{\tilde{Y}\tilde{Y}} = \sigma^2_{\tilde{y}_1} I$, and $R_{X^qX^q} = R_{XX} + \Delta R = R_{XX} + \sigma^2_{\tilde{y}_1} I$.
To optimize L, one should consider $\min_L \left(\det\left[diag(R_{YY})\right]\right)$, with this time $R_{YY} = L R_{XX} L^T + \sigma^2_{\tilde{y}_1} \overline{L}\,\overline{L}^T$. One can show that the resolution of the normal equations leads to the following expression for the coding gain $G_L^{(1)}$, taking into account the perturbations up to first order

$$G_L^{(1)} \approx \left(\frac{\det\left[diag(R_{XX})\right]}{\det\left[diag(L R_{XX} L^T + \sigma^2_{\tilde{y}_1} \overline{L L}^T)\right]}\right)^{\frac{1}{N}} \quad (9)$$

with $L R_{XX} L^T = D$ and $\sigma^2_{\tilde{y}_1} = c\, 2^{-2R}(\det D)^{\frac{1}{N}}$ where D is the diagonal matrix of the non perturbed prediction error variances, and L and $\overline{L}$ are also non perturbed quantities. This expression is established under the high resolution assumption ($\sigma^2_{\tilde{y}_1} I$ is small in comparison with R_{XX}). The term $\sigma^2_{\tilde{y}_1} \overline{L L}^T$ shows that the prediction error variance of the current sample y_i is increased by the quantization noise of the previous samples, filtered by the energy of $L_{i,1:i-1}$. Another interesting expression of $G_L^{(1)}$

is

$$G^{(1)}L \approx G_L^{(0)}\left(1-\frac{\sigma^2_{\tilde{y}_1}}{N}\left(\Sigma_{i=1}^N\frac{1}{\lambda_i}-\Sigma_{i=1}^N\frac{1}{\sigma^2_{y_i}}\right)\right) \quad (10)$$

where $G_L^{(0)}$ is the coding gain in the ideal case, $\{\lambda_i\}$ the eigenvalues of the autocorrelation matrix of X, and $\sigma^2_{\tilde{y}_1}$ the quantization noise in the ideal case, assumed to be white. Thus, maximizing the coding gain entails maximizing the sum of the inverses of the prediction error variance. Whereas the coding gain in the ideal case is invariant by permutation, there is in the backward adapted causal transform coding scheme an optimal ordering of the components of the X_i.

3.2 Unitary Approach

The authors in [4] have studied transform coding schemes based on a backward adapted KLT. Under some assumptions, (use of a dither, no estimation noise) they have established the convergence of the sequence of successive estimates of T (based on the successive estimates of $R_{X^qX^q}$) to an optimal KLT T of R_{XX}. Let T denote a KLT of R_{XX}. Then $T(R_{XX}+\sigma^2_{\tilde{y}_1}I)T^T=\Lambda+\sigma^2_{\tilde{y}_1}I$, and T is also a KLT of $R_{XX}+\sigma^2_{\tilde{y}_1}I$. Thus, the perturbation term $\sigma^2_{\tilde{y}_1}I$ on R_{XX} does not change the backward adapted transformation, and

$$G^{(1)}_{KLT}=G^{(0)}_{KLT}. \quad (11)$$

The previous expression and (9) are asymptotic gains, that is, obtained with an infinite number of datas. An interesting question is that of the effects of the estimation noise on the computation of the transformations. This is the point of the following section.

4 Estimation Noise

Consider a vectorial process whose samples X_i are *i.i.d.*. The typical estimate of R is $\hat{R}=R+\Delta R=\frac{1}{k}\sum_{i=1}^k X_iX_i^T$. In this case, it is easy to see that ΔR is symmetric. $\Delta R=\hat{R}-R$ is the perturbating term occuring when a finite amount of k samples is used to compute the estimate $\hat{R}$ of the covariance matrix R (in the following, R_{XX} will be denoted by R). Note that no quantization effect is taken into account here. The question we want to address is : how does ΔR affect the expected coding gains G^0 of the two transformations, and which approach should be preferable ?

4.1 Perturbation of the LDU

The estimate of the prediction matrix based on $R+\Delta R$ is $L+\Delta\overline{L}$, with

$$(L+\Delta\overline{L})(R+\Delta R)(L+\Delta\overline{L})^T=R_{YY}+\Delta(R_{YY}), \quad (12)$$

where $\Delta(R_{YY})$ is a diagonal matrix, $\Delta\overline{L}$ is strictly lower triangular, and L, R and R_{YY} are non perturbed quantities. Thus using (5), we have up to the first order of the perturbation

$$L\Delta RL^T+\Delta\overline{L}L^{-1}R_{YY}+R_{YY}L^{-T}\Delta\overline{L}^T=\Delta(R_{YY}). \quad (13)$$

Since $\Delta\overline{L}L^{-1}R_{YY}$ and $R_{YY}L^{-T}\Delta\overline{L}^T$ are strictly lower triangular, we get

$$\left\{\begin{array}{rcl}\Delta(R_{YY}) & = & diag\{L\Delta RL^T\}\\ \Delta\overline{L}L^{-1}R_{YY}+\triangleright(L\Delta RL^T) & = & 0,\end{array}\right. \quad (14)$$

where $\triangleright(.)$ (resp. $\triangleleft(.)$) denotes the strictly lower triangular (resp. upper) matrix made with the strictly lower triangular part of (.). Now, the perturbed matrix $L+\Delta\overline{L}$ is applied to the signal to be coded, and we have

$$(L+\Delta\overline{L})R(L+\Delta\overline{L})^T=R_{YY}+\Delta(R_{YY})', \quad (15)$$

where $\Delta(R_{YY})'$ is the perturbation matrix of R_{YY}. Then, up to the first order of the perturbation

$$diag\Delta(R_{YY})'=diag\{\Delta\overline{L}L^{-1}R_{YY}+R_{YY}L^{-T}\Delta\overline{L}^T\}=0. \quad (16)$$

The diagonal matrix of the perturbation of the prediction error variances remains unchanged. Up to the second order, one finds

$$\begin{array}{rcl}diag\Delta(R_{YY})' & = & diag\{\Delta\overline{L}R\Delta\overline{L}^T\}\\ & = & diag\{\Delta\overline{L}L^{-1}R_{YY}R_{YY}^{-1}R_{YY}L^{-T}\Delta\overline{L}^T\},\end{array} \quad (17)$$

and using (14),

$$\begin{array}{rcl}diag\Delta(R_{YY})' & = & diag\{-\triangleright(L\Delta RL^T)R_{YY}^{-1}(-\triangleleft(L\Delta RL^T))\}\\ & = & diag\{\triangleright(L\Delta RL^TR_{YY}^{-\frac{1}{2}})(\triangleleft(R_{YY}^{-\frac{1}{2}})L\Delta RL^T)\}.\end{array} \quad (18)$$

Let $\delta_i^{'}$ be the i-th element of $\Delta(R_{YY})^{'}$, then

$$\begin{array}{rcl}\delta_i^{'} & = & \{L\Delta RL^TR_{YY}^{-\frac{1}{2}}\}_{i,1:i-1}\{L\Delta RL^TR_{YY}^{-\frac{1}{2}}\}^T_{i,1:i-1}\\ & = & \|(L\Delta RL^TR_{YY}^{-\frac{1}{2}})_{i,1:i-1}\|^2.\end{array} \quad (19)$$

Now, the coding gain $G^{(1)}_{L+\Delta\overline{L}}$ obtained with the transformation $L+\Delta\overline{L}$ is

$$\begin{array}{rcl}G^{(1)}_{L+\Delta\overline{L}} & = & \left(\frac{\prod_{i=1}^N\sigma^2_{x_i}}{\prod_{i=1}^N(\sigma^2_{y_i}+\delta_i^{'})}\right)^{\frac{1}{N}}\\ & \approx & \left(\frac{\prod_{i=1}^N\sigma^2_{x_i}}{\prod_{i=1}^N\sigma^2_{y_i}}\right)^{\frac{1}{N}}\left(\frac{1}{\sum_{i=1}^N 1+\frac{\delta_i^{'}}{\sigma^2_{y_i}}}\right)^{\frac{1}{N}}\\ & \approx & G^{(0)}(1-\frac{1}{N}\sum_{i=1}^N(1+\frac{\delta_i^{'}}{\sigma^2_{y_i}})).\end{array} \quad (20)$$

With (19), we have

$$\sum_{i=1}^N\frac{\delta_i^{'}}{\sigma^2_{y_i}}=\sum_{i=1}^N\|(R_{YY}^{-\frac{1}{2}}L\Delta RL^TR_{YY}^{-\frac{1}{2}})_{i,1:i-1}\|^2, \quad (21)$$

that is, the sum of the lower off diagonal elements of $R_{YY}^{-\frac{1}{2}}L\Delta RL^TR_{YY}^{-\frac{1}{2}}$, which is symmetric. Hence, denoting by $\|A\|_F$ the Frobenius norm of the matrix A, and by tr the trace operator,

$$\begin{array}{l}\sum_{i=1}^N\frac{\delta_i^{'}}{\sigma^2_{y_i}}\\ =\frac{1}{2}\left[\|R_{YY}^{-\frac{1}{2}}L\Delta RL^TR_{YY}^{-\frac{1}{2}}\|_F^2-diag\{\|R_{YY}^{-\frac{1}{2}}L\Delta RL^TR_{YY}^{-\frac{1}{2}}\|_F^2\}\right]\\ =\frac{1}{2}\left[tr\{R_{YY}^{-\frac{1}{2}}L\Delta RL^TR_{YY}^{-\frac{1}{2}}R_{YY}^{-\frac{1}{2}}L\Delta RL^TR_{YY}^{-\frac{1}{2}}\}-\|R_{YY}^{-1}\Delta(R_{YY})\|_F^2\right]\\ =\frac{1}{2}\left[tr\{\Delta RR^{-1}\Delta RR^{-1}\}-\|R_{YY}^{-1}\Delta(R_{YY})\|_F^2\right]\end{array} \quad (22)$$

We get finally the following gain when the estimation noise is taken into account

$$G^{(1)}_{L+\Delta\overline{L}} \approx G^{(0)}\left(1-\frac{1}{2N}\left[tr\Delta RR^{-1}\Delta RR^{-1}\right]-\|R^{-1}_{YY}\Delta(R_{YY})\|^2_F\right) \quad (23)$$

The expectations of the two terms in (23) are now computed separately.
First term in (23): Let $K = R^{-\frac{1}{2}}\Delta RR^{-\frac{T}{2}}$, then this term becomes $tr\{KK\}$ which is also $\|vec(K)\|^2 = vec^T(K)vec(K)$. By using the following property

$$vec(K) = (R^{-\frac{1}{2}} \otimes R^{-\frac{1}{2}})vec(\Delta R), \quad (24)$$

where $\otimes$ denotes the Kronecker product, one can show that for i.i.d. vectors, $vec(K) \sim \mathcal{N}(\mathbf{0}, \overline{R})$, with

$$\overline{R} = Evec(K)vec^T(K) = \frac{1}{k}I_{N^2}. \quad (25)$$

Thus, taking expectation of the first term yields

$$Etr\{\Delta RR^{-1}\Delta RR^{-1}\} = tr\{Evec(K)vec^T(K)\} = \frac{N^2}{k}. \quad (26)$$

Expectation of the second term in (23) :

$$E\|R^{-1}_{YY}\Delta(R_{YY})\|^2_F = NEtr\{\Delta RR^{-1}\Delta RR^{-1}\}_{N=1} = \frac{N}{k}. \quad (27)$$

Finally, the expectation of the perturbed gain is

$$EG^{(1)}_{L+\Delta\overline{L}} = G^{(0)}\left(1-\frac{1}{2N}\frac{N(N-1)}{k}\right) = G^{(0)}(1-\frac{N-1}{2k}). \quad (28)$$

The expected coding gain in the presence of estimation noise is independent of R.

4.2 Perturbation of the KLT

A similar analysis of the perturbation can be lead in the unitary case. First, the estimated transformation $T+\Delta T$ is such that

$$(T+\Delta T)(R+\Delta R)(T+\Delta T)^T = \Lambda + \Delta\Lambda, \quad (29)$$

where $\Delta\Lambda$ is a diagonal matrix. Note also that ΔTT^T is antisymmetric since

$$TT^T = I \Rightarrow \Delta TT^T + T\Delta T^T = 0. \quad (30)$$

Similarily to (15), we have

$$(T+\Delta T)R(T+\Delta T)^T = \Lambda + \Delta\Lambda', \quad (31)$$

where $\Delta\Lambda'$ is the perturbation matrix due to the use of $T+\Delta T$ in the coding procedure. Then, one shows from (31) and (30) that

$$diag\{\Delta\Lambda'\} = diag\{\Delta T^T R\Delta T\}. \quad (32)$$

Also, the perturbed coding gain $G^{(1)}_{T+\Delta T}$ can be expressed as

$$G^{(1)}_{T+\Delta T} \approx G^{(0)}\left(1-\frac{1}{N}\sum_{i=1}^{N}(1+\frac{\delta\lambda'_i}{\lambda_i})\right) \quad (33)$$

where the $\delta\lambda'_i$ denote the diagonal elements of $\Delta\Lambda'$. Thus from (32)

$$G^{(1)}_{T+\Delta T} \approx G^{(0)}\left(1-\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\lambda_i}\Delta T_i^T R\Delta T_i\right), \quad (34)$$

and by taking the expectation of (34), we get

$$E\ G^{(1)}_{T+\Delta T} \approx G^{(0)}\left(1-\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\lambda_i}tr\{RE\ \Delta T_i\Delta T_i^T\}\right). \quad (35)$$

Using the following classical result in perturbation theory of matrices ([1])

$$E\ \Delta T_i\Delta T_i^T = \frac{\lambda_i}{k}\sum_{j\neq i}\frac{\lambda_i}{(\lambda_j-\lambda_i)^2}T_iT_i^T, \quad (36)$$

and the expectation of the perturbed gain is

$$E\ G^{(1)}_{T+\Delta T} \approx G^{(0)}\left(1-\frac{1}{k}\sum_{i=1}^{N}\sum_{j\neq i}\frac{\lambda_i\lambda_j}{(\lambda_j-\lambda_i)^2}\right). \quad (37)$$

4.3 Discussion

On the one hand, the expected coding gain in (28) is independent of the signal statistics and the bias is linear in the dimension of the problem (N). The number of measurements k should be several times bigger than the dimension of the problem in order to be close to optimality.
On the other hand, the expected coding gain for the KLT depends on the signal, and the degradation term in (37) involves this time $N(N-1)$ terms. Nevertheless, it is difficult in general to determine which approach yields the best coding gain. One can see that $EG^{(1)}_{T+\Delta T}$ can for some R be better and for some other worse. Let us take for example a very ill-conditoned matrix R whose eigenvalues are $1, \epsilon, \epsilon^2, ..., \epsilon^{N-1}$, and let ϵ go to zero. Then each term of the double sum in (37) will be arbitrarily small and $EG^{(1)}_{T+\Delta T}$ close to $G^{(0)}$. If now two consecutive eigenvalues are very close, $EG^{(1)}_{T+\Delta T}$ can decrease arbitrarily. Such variations do not occur with the causal LDU approach.

References

[1] S. Haykin, Editor, *Advances in Spectrum Analysis and Array Processing, vol. II*, Advanced Reference Series, Prenctice Hall.

[2] D. Mary, Dirk T.M. Slock, "Codage DPCM Vectoriel et Application au Codage de la Parole en Bande Élargie," in *CORESA 2000*, Poitiers, France, October 2000.

[3] S.-M. Phoong, Y.-P. Lin, "Prediction-Based Lower Triangular Transform," *IEEE trans. on Sig. Proc.*, vol. 48, no. 7, July 2000.

[4] V. Goyal, J. Zhuang, M. Vetterli, "Transform Coding with Backward Adaptive Updates," *IEEE trans. on Inf. Theory*, vol. 46, no. 4, July 2000.